(* COUNTS *)

30

Unordered responses, many paired

		Exh	Page
Introduction	30-1		
		A	
		1	30-4
30A. Two responses, both unordered	30-2	B	
		2	30-9/10
30B. Unaffectedness -- the reference standard	30-6	C	
		3	30-14
30C. Pictures for multiplicative fits	30-12	4	30-16
- fitted only	30-13	5	30-17
- observed and fitted	30-15	6	30-19/20
- an alternate scheme	30-20	7	30-21
		D	
30D. Here come the logs!	30-21	8	30-25/27
- logs or square roots?	30-24	9	30-30
- logs are not all alike	30-28	E	
		10	30-32
30E. Circumstance and response, both unordered	30-31	11	30-33
- ordering can sometimes be extracted	30-40	12	30-35/36
		13	30-37
30F. Ordered circumstance and unordered response	30-42	14	30-38
- a second ordering?	30-46	15	30-41
		F	
**********		16	30-43
		17	30-44
30H. How far have we come?	30-48	18	30-47
		H	
**********		K	
		19	30-51/52
30K. A further example	30-50	P	
		20	30-54
**********		21	30-55
		22	30-56/57
30P. Problems for chapter 30	30-53		

30

Unordered Responses

Sometimes a two-way table of counts conceals a response whose versions are unordered. In such a case, we cannot immediately use the powerful techniques that tend to squeeze ordered responses dry.

If one factor is an unordered circumstance, while the other is an unordered response, we can do little beyond asking how the fractions of each response compare from one circumstance to another. In doing this, of course, we are free to use all the techniques we have learned or can invent.

One difficulty that we always have to face is how to arrange our results -- either numbers or graphics -- in an unordered way. Showing absence of order in such a way that the elements can be effectively compared is not easy.

If one factor is an ordered circumstance, the other an unordered response, we can do rather better. If there are not too many versions of the response, we can look at the prevalence of each as a function of the circumstances. Here the order in the circumstance helps considerably -- we have a chance of seeing trends or isolated irregularities.

As a result we are likely -- even for an unordered circumstance -- to focus on changes in prevalence of an individual response as a function of circumstance.

We ought really, it would seem, to look at unordered responses as a bundle of counted fractions -- one for each version of the response -- all together summing up to unity -- we shall do this in chapter 30x. We may find, however, that treating them as counts rather than counted fractions will either work well enough or, perhaps, even work better.

When we come to two-way tables concealing a pair of unordered responses, the same attitude would lead us to say that we ought to consider them as a two-way bundle of 2-by-2 tables. We will try this when we reach chapter 30x.

Alternatively, we could merely analyze the logs of the counts. This might be either better or worse.

It will not always be clear how we ought to look at the factors in a two-way table of counts. We ought, therefore, to regard trying two or more analyses as quite usual.

30A. Two responses, both unordered

Two-way tables both of whose factors are responses look like -- and can sometimes by analyzed like -- two-way tables both of whose factors are circumstances. The thinking involved, however, cannot be the same -- the types of questions it is natural to ask are likely to be quite different.

An example is in order. A classic example will serve us well, one used by Yule in his An Introduction to the Theory of Statistics (first edition 1910, tenth edition, revised, 1936) and credited to Ammon, Zur Anthropologie der Badener, concerning the hair and eye color of 6800 males in Baden. The data is given in panel A of exhibit 1.

This is a two-way table of counts. We are going to begin by analyzing it in a way that is surprisingly similar to the way we analyzed the data of exhibit 7 of chapter 24, repeated for comparison in panel B of exhibit 1. There the counts were white grubs in turf samples, arranged by lead arsenate treatment and block. Here the counts are of people, arranged by eye color and hair color. There is a major difference -- one that we need to understand clearly.

In the earlier example, blocks of land were there, certain parts of them were made into plots and treated with different amounts of lead arsenate. These were circumstances. In due course the various plots supported different populations of white grubs. Samples were taken and counted. These counts were responses.

The eight-year gap between layout and treatment on the one hand and sampling and counting on the other makes the distinction between circumstance and response quite clear.

exhibit 1 of chapter 30 -- comparison

Two very different kinds of two-way tables

A) AMMON on MEN NATIVE to BADEN

Eye color	Hair color				
	Fair	Brown	Black	Red	(sum)
Blue	1768	807	189	47	(2811)
Gray or Green	946	1387	746	53	(3132)
Brown	115	438	288	16	(857)
(sum)	(2829)	(2632)	(1223)	(116)	6800

See following exhibit

B) NEISWANDER on WHITE GRUBS NATIVE to TURF

Block	Lead arsenate 8 years earlier					
	0	5	10	20	30	(sum)
1	16	5	1	0	1	(23)
2	10	3	1	1	0	(15)
3	10	6	0	2	1	(19)
4	7	6	0	0	0	(13)
5	14	5	6	4	1	(30)
6	26	10	6	2	0	(44)
7	24	14	6	1	0	(45)
8	20	14	4	3	1	(42)
(sum)	(127)	(63)	(24)	(13)	(4)	231

See exhibit 7 of chapter 24.

This sort of delay--long or short-- is not always a part of the distinction between circumstances and response, but it does help to make things clear when it is present.

The table of counts on the Badeners is quite a different matter. When Ammon made his study, certain people lived in Baden. He might--at least in principle--have examined all of them. He stopped, in practice, with a sample. This was reasonable -- 6800 is a fairly large sample. Each person had many characteristics. Two, at least, were measured by Ammon--hair color and eye color. These are of the nature of responses--not to some planned treatment, but to the things that had shaped the Baden population: early inhabitants, marriage practices, fertility, immigration, emigration. In looking at Ammon's table we are trying to say how these responses are behaving.

This question splits naturally into two questions:

- how do the individual responses behave?
- how do they relate to one another?

The latter question tends to fall apart into a general question and a particular one:

- what is the simplest sort of joint behavior?
- do these particular responses seem to deviate from simplicity?

Our next task is to clarify this general question. Before we begin, let us turn back to the two examples again

and point out another way in which their differences is shown. Think about the column and row sums. In the Baden example, these sums are individually meaningful. They tell us how many of the 6800 Badeners had a given hair color or a given eye color. It would have made sense to have either of these sets of numbers alone.

For the white grubs, the situation is different. The row sums, for instance, tell us about total numbers of grubs in samples from 5 plots treated quite differently. They are--or could be--useful summaries, but their meaning is not in themselves--it comes from the meanings of the individual counts they summarize. The same is true--somewhat less vigorously became these blocks do not vary as much as they might--for the column sums.

Asking ourselves how we ought to think about stripe sums often helps us to become clearer about the factor summed over--is it a circumstance or a response?

If we think about either response alone, we recognize that its counts are really numerators of counted fractions. So long as these fractions do not get too large, we are probably safe enough in treating them like counts.

30B. Unaffectedness -- the reference standard

The simplest relation of hair color to eye color is an "absence of relationship" . This would be shown if the distribution over eye color of, say, 10,000 red-haired men were the same as the distributions over eye color of 10,000 men of each of the other hair colors.

Expressing this standard situation for the entries of a single complete table gives this relation:

(fraction of cases in cell)

=

(fraction in row) x (fraction in column)

If the entries for some cells in the table are absent, we have to define all these fractions carefully. In general, however, we will have a relation of the form

(count in cell) = (a number for row) x (a number for column)

Multiplication is a basic element in our standard of unaffectedness.

Let us look at an extreme example, one involving two rare happenings that can--but need not--occur together. If one, A, happens 1% of the time, the other, B, 2%, and we have examined exactly 10,000 cases, the following 2-by-2 table is generally agreed to represent a complete lack of relationship of each event to the other.

	not A	A	(totals)
not B	9702	98	(9800)
B	198	2	(200)
(totals)	(9900)	(100)	((10,000))

Knowing about A tells us nothing about B--whether we have A or not A, exactly two cases in 100 are B's. B tells us--equally--nothing about A. The characteristic of this null table is its multiplicative character--multiplicative not additive.

Let us look at one cell in detail. The fraction of cases in the 'A' - 'not B' cell, .0098 in the example, is equal to the result of multiplying the fraction in the 'A' column, .01 in the example, by the fraction in the 'not B' row, .98 in the example. (We can repeat this for the other three cells with similar results.)

This acceptance of multiplicativeness as a standard of unaffectedness is essential to how we think about the world. Changing it would be almost inconceivable.

Let us return to the Badeners and put this standard to work. Exhibit 2 gives the first set of calculations. Starting from the marginal totals, we calculate fitted counts.

As shown by the example at the foot of the exhibit, the working arithmetic is very simple. The reason behind it is almost equally simple, namely

$$\text{fraction with blue eyes} = \frac{2811}{6800}$$

$$\text{fraction with fair hair} = \frac{2829}{6800}$$

$$\text{product of these} = \frac{(2811)\ (2829)}{(6800)\ (6800)}$$

The last of these is thus the fitted fraction for the extreme NW cell. To change this from a fraction to a count, we need to multiply it by 6800. This cancels one factor of 6800 in the denominator and gives the form shown in panel G.

In finding a square root corresponding to the observed counts we have used double roots. As was suggested in chapter 24,

exhibit 2 of chapter 30 -- Hair and eyes

Ammon's data on hair and eye-color from 6800 males in Baden

Eye Color	Hair Color				(sum)
	Fair	Brown	Black	Red	
A) OBSERVED COUNTS					
Blue	1768	807	189	47	(2811)
Gray or Green	946	1387	746	53	(3132)
Brown	115	438	288	16	(857)
(sum)	(2829)	(2632)	(1223)	(116)	6800
B) FITTED COUNTS -- from marginal totals, see example in panel G					
Blue	1169	1088	506	48.0	(2811)
Gray or Green	1303	1212	563	53.4	(3131.4)
Brown	357	332	154	14.6	(857.6)
(sum)	(2829)	(2632)	(1223)	(116.0)	((6800))
C) DOUBLE ROOTS for OBSERVED -- $(4(\text{obs}) + 2)^{\frac{1}{2}}$					
Blue	84.1	56.7	27.5	13.8	
Gray or Green	61.5	74.5	54.5	14.6	
Brown	21.5	42.0	34.0	8.1	
D) DOUBLE ROOTS for FITTED -- $(4(\text{fit}) + 1)^{\frac{1}{2}}$					
Blue	68.4	66.0	45.0	13.9	
Gray or Green	72.4	69.5	47.6	14.6	
Brown	37.8	36.4	24.8	7.7	
E) RESIDUALS -- difference of last two					
Blue	15.7	-9.3	-17.5	-0.1	
Gray or Green	-10.9	5.0	6.9	0.0	
Brown	-16.3	5.6	9.2	0.4	

F) SOURCE Ammon, Zur Anthropologie der Badener.

G) EXAMPLES $1169 = \frac{(2811)\ (2829)}{(6800)}$

$$84.1 = \sqrt{4(1768) + 2} = (4(1768) + 2)^{\frac{1}{2}}$$

(Cont'd)

exhibit 2 of chapter 30 (Cont'd)

$$68.4 = \sqrt{4(1169) + 1} = (4(1169) + 1)^{\frac{1}{2}}$$

$$15.7 = 84.1 - 68.4$$

H) ALSO USED by:

M. G. Kendall 1945, The Advanced Theory of Statistics volume 1, page 300 (2nd edition) Table 12.4.

G. Udny Yule 1910. An Introduction to the Theory of Statistics (1st edition 1910, 10th edition 1936) Table II on page 61.

we have used $\sqrt{2 + 4(\text{observed})}$ and $\sqrt{1 + 4(\text{fitted})}$. We shall not try to give an adequate explanation for this distinction since it would take us far outside the scope of the present account.

((For those who must have a hint, it will have to be enough to know that the reason for such a distinction between how we square-root an observed value and how we square-root a fitted value is closely linked to the downward curvature of the square root. If we turn th exhibit 19 of chapter 3, for instance, and consider where the curve for the (matched) square root will have to fall, we see that it must fall below the straight line $y = x$ at both ends. Thus it will be generally somewhat lower. We compensate for some of this when we use "+1" for fitted as against "+2" for observed.))

So much for the arithmetic and the reasons for its use. What about the results of the analysis? In particular, what about the residuals? We see numbers of varying size among them. Those for red hair are small, both relatively to the others and relatively to our previous experience. We have seen, in other examples where double roots were used, many residuals of size between 0 and 1, and rather fewer of sizes between 1 and 2. Instances of values larger than 2 -- or perhaps 3 -- that proved -- or seemed likely -- to be accidental

were quite rare. Values like 5, 6 and 9 are thus very large while values like 15, 16 and 17 are enormous. Since residuals of these sizes appear, we are almost forced to believe that, in Baden, eye color and hair color ARE associated.

Especially in tables with more rows, or more columns, or both, the fitting procedure underlying the analysis we have just done has both advantages and disadvantages. In particular:

- if each cell contains a rather well-behaved count, then the marginal totals are both natural and effective to use.
- if one (or a few) cells is unusual enough (usually by being unduly large) it can dominate the corresponding marginal totals, and thus distort the fit.

30C. Pictures for multiplicative fits

We would like to have a picture of both observed and fitted for counts involving two unordered factors, one that would be as responsive to the general principles set out in chapters 25 and 26 as is reasonable. In particular, we would like, if it is not too difficult, to have:

- areas proportional to total count.
- visual differences roughly proportional to differences in square roots.

If we can also expose the relative sizes of the one-way counts -- the stripe sums of the two-way counts -- this will be a valued bonus.

* fitted only *

If we want to show the fitted counts easily we have only to:

- write the fit as a product of row depths by column widths, and
- draw geometrical stripes with depths and widths proportional to those just written down. Equal spacing between adjacent stripes seems to work quite well.

To fix depths and widths we make some simple choice, depending on how the multiplicative fit was found. In the hair-and-eye-color example, where we found the fit for "Blue eyes" and "Fair hair" as

$$1169 = \frac{(2811)(2829)}{6800} = \frac{2811}{\sqrt{6800}} \cdot \frac{2829}{\sqrt{6800}} = (34.1)(34.3)$$

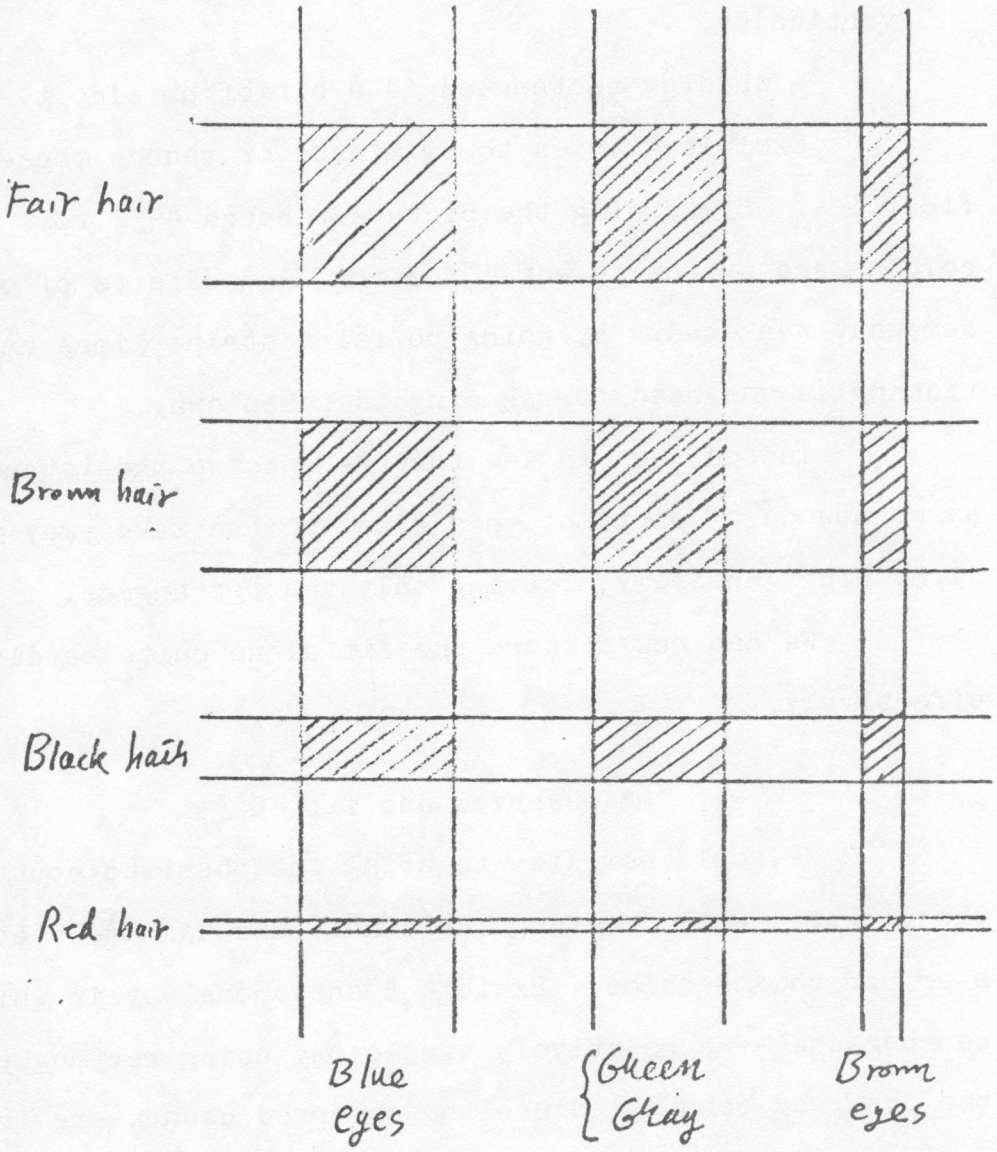
the natural choice is a depth of 34.1 or "Blue eyes" and a width of 34.3 for "Fair hair".

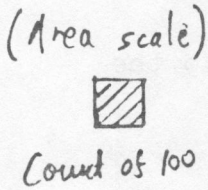
Exhibit 3 shows the result, with the meaningful rectangules shaded in. As a picture of the fit, this is effective, but has some disadvantages, including:

- it is hard to treat rows and column symmetrically. For any specific purpose it is easier to look at one rather than the other.

exhibit 3 of chapter 30 -- hair and eyes

The multiplicative fit to data
for 6800 inhabitants of Baden portrayed
(see exhibits 1 and 2)





- our versions are supposed to be unordered -- yet we must put them in some order. The order we choose matters most when they are arranged either horizontally or vertically.

- shading rectangles is a bit of an effort.

Exhibit 4 tries to eliminate or reduce these difficulties. By turning the picture through 45°; rows and columns are put on an equal footing, and effects of order are somewhat weakened. By going to faint stripe lines and heavy rectangles the need for shading is taken away.

Indeed, we can see that -- once we understand as much as we now do of plots of this kind, we can take away the light lines entirely, leaving only the rectangles.

We can now picture the fit alone quite easily and effectively.

* observed and fitted *

We would now like to bring the observed counts into the picture, while meeting the conditions laid down at the start of this section. Exhibit 5 shows one way in which this can be done -- a relatively simple way using rectangles for the observed counts. Since the observed counts are irregular in size, they deserve to be the solid rectangles.

The solid rectangles have area proportional to the observed counts. They are so arranged that

observed > fitted

exhibit 4 of chapter 30 -- hair and eyes

The picture of exhibit 3
rotated through 45° and modified in technique
(see exhibits 1 to 3)

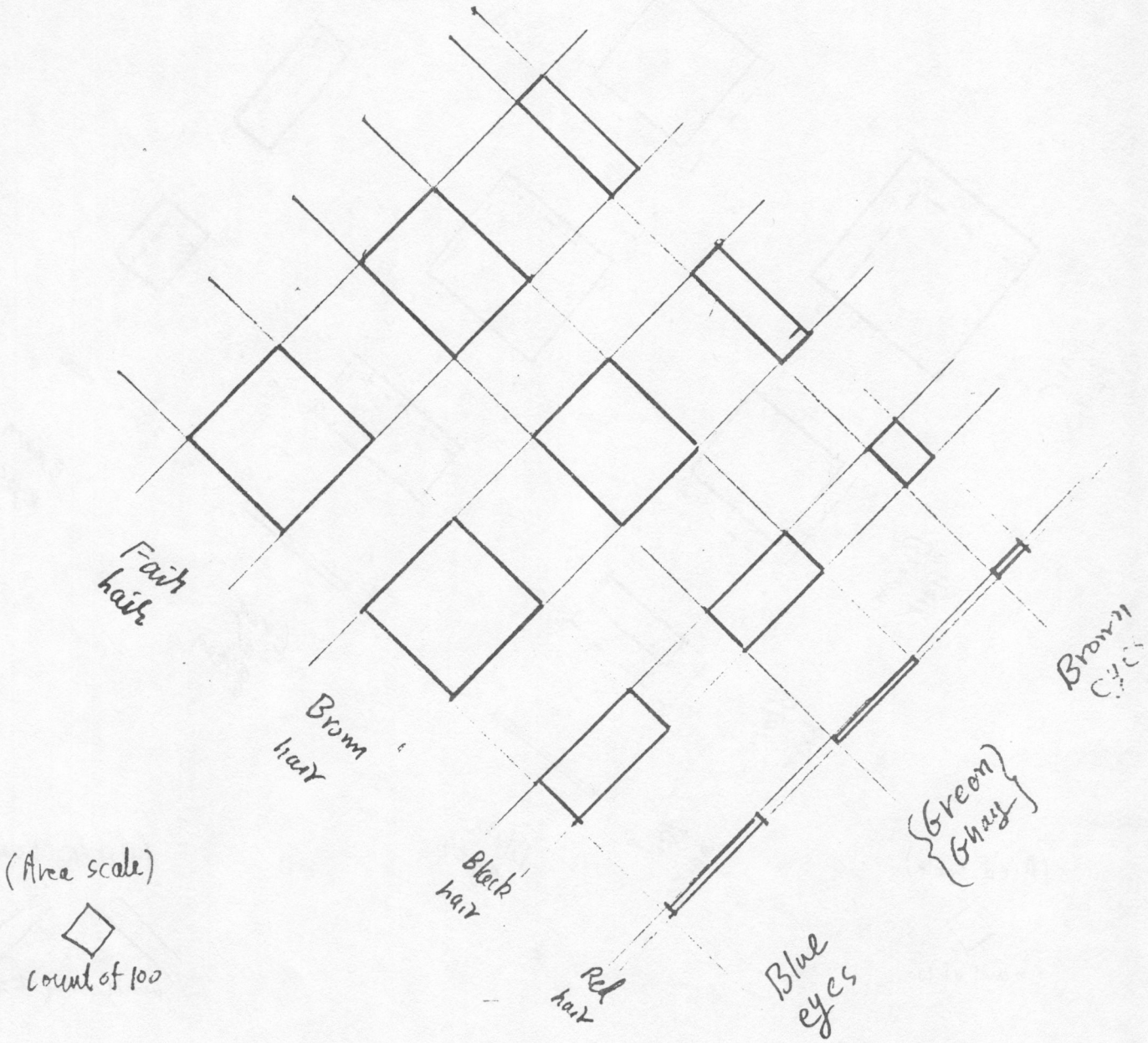
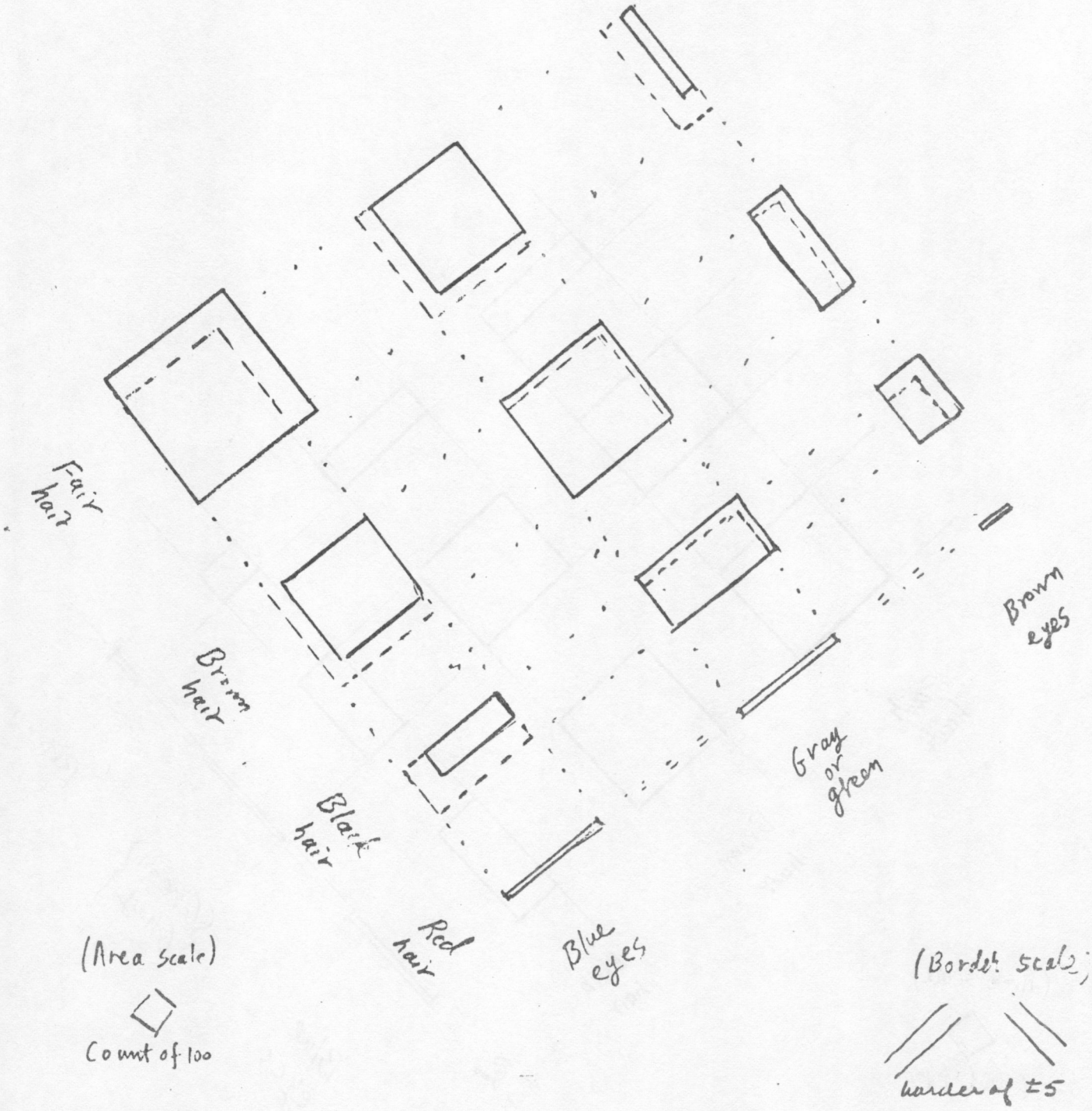


exhibit 5 of chapter 30 -- hair and eyes

Observed (solid rectangles) and fitted (dashed rectangles) counts for Ammon's data on inhabitants of Baden (area scale only, see exhibit 6 for calculations)



is shown by an equal failure to reach the two lower sides of such a rectangle.

If depth and width are the sides of a fitted rectangle in units such that

$$(\text{depth})\times(\text{width}) = \text{fitted count}$$

and "border" is the width by which depth and width are both changed -- increased if border is "+", decreased if it is "-" -- make the observed rectangle then

$$(\text{depth})\times(\text{width}) + (\text{depth} + \text{width})\times(\text{border}) + (\text{border})$$

equals

fitted count + residual count.

Subtracting (depth) + (width) from both sides gives

$$(\text{depth} + \text{width} + \text{border})(\text{border}) = \text{residual count}$$

whence

$$\text{border} = \frac{\text{residual count}}{\text{depth} + \text{width} + \text{border}}$$

a simple relation easily solved by successive approximation.

Exhibit 6 has the necessary arithmetic. It is easy, but not quite short.

exhibit 6 of chapter 30 -- hair and eyes

The calculations needed for exhibit 5
(see also exhibits 1 to 4)

A) The RESIDUAL COUNTS -- as observed MINUS fitted from exhibit 1

	Hair color			
	Fair	Brown	Black	Red
Blue eyes	599	-289	-317	-1
Gray or green	-357	175	183	-0h
Brown eyes	-242	106	134	1h

Examples: 599 = 1768 - 1169, -357 = 946 - 1303.

B) The values of DEPTH, WIDTH and DEPTH + WIDTH

	34.3	31.9	14.8h	1.4
34.1	68.4	66.0	49.9h	35.2
38.0	72.3	69.9	52.8h	39.4
10.4	44.7	42.3	25.2h	11.8

Examples: $34.1 = 2811/\sqrt{6800}$, $38.0 = 3132/\sqrt{6800}$, ...
$34.3 = 2829/\sqrt{6800}$, $31.9 = 2632/\sqrt{6800}$, ...
68.4 = 34.3 + 34.1, 72.3 = 34.3 + 38.0, ...

C) The values of BORDER

Blue eyes	7.8	-4.6	-7.7	-.0
Gray or Green	-5.3	2.4	3.3	-.0
Brown eyes	-6.2	2.4	4.5	.1

Example (successive stages in calculating 7.8):

trial border	depth + width + border	599/(depth + width + border
0	68.4 + 0 = 68.4	599/68.4 = 8.8
8.8	68.4 + 8.8 = 77.2	599/77.2 = 7.8
7.8	68.4 + 7.8 = 76.2	599/76.2 = 7.8, stops

(A similar set of steps can be taken for each other cell. Short-cutting by judgment can decrease the number of steps needed. With a slide-rule the process goes quite easily and rapidly.)

Checks: 72.3 + (-5.3) = 67.0, -357/67.0 = -5.3
44.7 + (-6.2) = 38.5, -242/38.5 = -6.2,

(cont'd.)

* an alternate scheme *

Some of us will like to shorten the arithmetic in any reasonable way. Exhibit 7 shows one way in which this can be done. As shown, the first step in calculating an exact border

$$\text{first border} = \frac{\text{residual count}}{\text{depth + width}}$$

is a good enough approximation for pictorial purposes.

As panel B of exhibit 7 also shows for this set of data, both "border" and "first border" behave rather like half the difference between observed and fitted double roots. Indeed, the agreement is quite close for not too large residuals. This offers another approximation good enough for pictorial purposes. We shall all feel free to use either approximation whenever it seems convenient.

30D. Here come the logs!

There is a real place for a fitting procedure based on marginal totals; there is another, equally real place for a fitting procedure that is more resistant to one -- or a few -- erratic cells. When we start to cast around for such a procedure

exhibit 7 of chapter 30 -- hair and eyes

Some approximations to "border"

A) Values of FIRST BORDER -- (residual) / (depth + width)

Blue eyes	8.8	-4.3	-7.1	-.0
Gray or green	-4.9	2.5	3.5	-.0
Brown eyes	-5.6	2.5	5.3	.1

Examples: 8.8 = 599/68.4, -4.9 = -357/72.3, ...

B) COMPARISON of BORDER with RESIDUALS for DOUBLED ROOT and FIRST BORDER

Hair	Eyes	Border	(1/2) Residual	First Border
Fair	Blue	7.8	7.8	8.8
Black	Brown	4.5	4.6	5.3
Black	G or G	3.3	3.4	3.5
Brown	Brown	2.4	2.8	2.5
Brown	G or G	2.4	2.5	2.5
Red	Brown	.1	.2	.1
Red	G or G	.0	.0	.0
Red	Blue	.0	.0	.0
Brown	Blue	-4.6	-4.6	-4.3
Fair	G or G	-5.3	-5.4	-4.9
Fair	Brown	-6.2	-8.2	-5.6
Black	Blue	-7.7	-8.7	-7.1

Note: "Border" from panel C of exhibit 6,
"first border" from panel A above,
"Residual" from panel E of exhibit 2.

NO PAGE 22

we should recall that we have already set up resistant procedures for analyzing two-way tables of responses -- procedures that depend on additive effects. It is natural to want to try to use the same approach here.

We have a multiplicative implied reference -- how can we make it an additive one? As always, the answer is "Take logs." If the cell counts fit a multiplicative model perfectly, their logs will fit an additive one equally perfectly! So long as there are no zero values plain logs will do. If there are zeros, we can give the logs a free start, using either log(count + 0.1) or log(count + 0.25).

We might even, perhaps, if there are not too many zero entries, just leave out any zero entry. This is rarely satisfactory, however, since we lose all trace of the fact that the count in that cell was small. Starting the logs is usually much better.

The tolerance for many difficulties of resistant two-way analyses is great, which helps us greatly. Not only can we live with a small fraction of missing values, or with a somewhat larger fraction of (more or less arbitrarily) modified logs, but resistant two-way analyses also do much to reduce trouble from the fact that cells with small counts tend to have less firmly fixed logarithms than do cells with large counts.

Once we combine their use with resistant two-way analysis, logs look reasonably attractive. Exhibit 8 sets out an analysis using both.

In order to show the effect of including relatively small cells, exhibit 8 first analyzes the data omitting the column for red hair, and then goes through all the calculations again including this column. The difference between the two analyses -- as panel D shows -- amounts to -.08 in three of the nine residuals that can be compared. The change in the one corresponding row effect is .06, while those in the other row effects and column fits at most ±.03. Clearly, including the small cells made little difference here.

A verbal summary of the information in the residuals runs roughly as follows:

- Fair hair and blue eyes go together about three times as often as indicated by the row-and-column fit.
- Black hair and blue eyes go together about 1/2 as often as would be computed for the row-and-column fit.
- Fair hair and brown eyes go together about 1/3 as often as would be computed from the row-and-column fit.

* logs or square roots? *

If we do not need to worry about the fluctuations and irregularities that belong to our specific body of data, a fully logarithmic analysis does better in getting at the

exhibit 8 of chapter 30 -- hair and eyes

Ammon on the anthropology of the Badeners
this time analyzed by logs

A) RAW COUNTS -- again, see exhibits 1 and 2

Eye color	Fair	Brown	Black	Red
Blue	1768	807	189	47
Gray or Green	946	1387	746	53
Brown	115	438	288	16

----------(analysis for 3 hair colors)----------

B1) LOG COUNTS -- omitting red

Blue	3.25	2.94	2.28
Gray or green	2.98	3.14	2.87
Brown	2.06	2.64	2.46

B2) TWO-WAY ANALYSIS -- by median differences, or by medians, unit = 0.01 in logs

				eff	fit
Blue	47	0	-39	3	280
Gray or green	0	0	0	23	300
Brown	-42	0	9	27	250
eff	-2	14	-13	000	277
fit	275	291	264	277	000

B3) RESIDUALS AS % -- undoing logarithms, rounding

Blue	300%	100%	40%
Gray or Green	100%	100%	100%
Brown	40%	100%	125%

B4) STRIPE ADJUSTMENTS AS (multiple) RATIOS -- undoing logarithms

Blue/G-or-G/Brown = 1.00/1.58/0.50

Fair/Brown/Black = 1.00/1.45/0.78

B5) EXAMPLES: 3.25 = log 1768.

3.25 = .47 + .00 + .00 + 2.78

.47 = log 3.0, 3.0 = 30.0%

.16 = log 1.58.

(cont'd.)

exhibit 8 of chapter 30 (Cont'd)

----------(analysis for 4 hair colors)----------

C1) LOG COUNTS -- including red hair

Eye Color	Hair Color			
	Fair	Brown	Black	Red
Blue	3.25	2.94	2.28	1.67
Green or Gray	2.98	3.14	2.87	1.72
Brown	2.06	2.64	2.46	1.20

C2) TWO-WAY ANALYSIS -- using median differences

					eff	fit
Blue	.39	-.08	-.47	.07	.09	2.56
Green or Gray	.00	.00	.00	.00	.21	2.68
Brown	-.14	.01	.10	-.01	-.30	2.17
eff	.30	.46	.19	-.96	2.41	0.00
fit	2.27	2.93	2.65	1.51	0.00	-2.47

C3) RESIDUALS AS % -- undoing logarithms, rounding

Blue	250%	85%	35%	120%
Green or Gray	100%	100%	100%	100%
Brown	40%	100%	130%	10%

C4) STRIPE ADJUSTMENTS AS (multiple) RATIOS -- undoing logarithms

Fair/Brown/Black/Red = 1.15/1.66/0.89/0.06

C5) EXAMPLES 3.25 = log 1768.

3.25 = .39 + .06 + .00 + 2.80.

.39 = log 2.5, 2.5 = 250%

.06 = log 1.15.

(Cont'd)

exhibit 8 of chapter 30 (Cont'd)

D) DIFFERENCE OF ELEMENTARY ANALYSES -- only differences of comparable numbers shown

	Hair Color					
	Fair	Brown	Black	Red	eff	fit
Blue	-.08	-.08	-.08	-	+.06	-
Green or Gray	.00	.00	.00	-	-.02	-
Brown	.01	.01	.01	-	-.03	-
eff	-	-	-	-	***	-
fit	+.02	+.02	+.02	-	-	***

*** Large, not expected to be otherwise.

essence of the matter. If fluctuations and irregularities had been so large -- as they often are -- that we would be lucky to have modest confidence in pointing out some cells as "high" and others as "low", a square root comparison -- after a logarithmic analysis -- may do better in getting to those few aspects of the matter that might then be reachable.

This contrast between the usefulnesses of square-root and logarithmic methods, when both are applicable, applies generally, although we did not emphasize this enough in chapter 24. Roots help us in dealing with irregularities and fluctuations. Logs help us once the data is massive enough to be decisive. (As a variety of long-tailed examples show, logs sometimes help us in dealing with greater irregularities and fluctuations.)

* logs are not all alike *

We need to be clear about how the use of logs for the Badeners differs from the use of logs for Neiswander's grub counts. With grubs we went to logs -- more specifically to log(count + free start) with a free start probably between 0.25 and 1 -- because this gave the simplest description of the data. In logs, row and column effects accounted for nearly everything that seemed to be going on with the grub counts. We were free to choose our form of expression -- we wanted to choose it to make description simple.

In dealing with the Badeners, a man who wants to use an elementary analysis feels forced to go to logs, whether or not the results are simple. Our attitude toward the world -- as summed up in our definition of unaffectedness -- demands that an analysis of paired responses into a SUM of terms should be an analysis of logs. If some other form of expression were simpler, we would not use it.

Consider the dreamed-up data of exhibit 9. If 1234 and ABC referred to circumstances, we would feel free to use any form of expression that would help. The analysis in terms of raw counts -- see panel B -- cannot be improved on. Under these circumstances we would seize it.

If 1234 and ABC referred to responses -- as in dealing with the Badeners -- we would insist on analyzing log counts. Reading panel D, we would have to say that row and column fits do not tell the whole story -- each response fails to be unaffected by the other. This is clearly the case. Compare columns 1 and 4:

- in column 1 there are 400/1800 = 22.2% A's.
- in column 4 there are 100/900 = 11.1% A's.

Clearly the ABC response is not unaffected by selection on the 1234 response.

If unaffectedness is our standard of reference, an elementary analysis of log counts is very much more to

exhibit 9 of chapter 30 -- casual example

A set of counted data -- and two analyses

A) The COUNTS -- obviously invented for the analysis

	1	2	3	4
A	400	300	200	100
B	600	500	400	300
C	800	700	600	500

B) ELEMENTARY ANALYSIS of RAW COUNT

	1	2	3	4	eff	fit
A	0	0	0	0	-200	250
B	0	0	0	0	0	450
C	0	0	0	0	200	650
eff	150	50	-5	-50	450	000
fit	600	500	400	300	000	-450

C) LOGS of COUNTS

	1	2	3	4
A	2.60	2.48	2.30	2.00
B	2.78	2.70	2.60	2.48
C	2.90	2.85	2.78	2.60

D) ELEMENTARY ANALYSIS of C

	1	2	3	4	eff	fit
A	.08	.04	-.02	-.12	-.12	2.47
B	-.03	-.03	-.01	.07	.07	2.66
C	-.05	-.02	.03	.05	.05	2.64
eff	.17	.09	-.03	-.23	2.59	0.00
fit	2.76	2.68	2.56	2.36	0.00	-2.59

the point than are elementary analysis of say, raw counts. This is generally true. In this particular example, if simple analysis is our aim, the opposite applies. There can be very real differences.

Not all elementary analyses of log counts are alike.

30E. Circumstance and response, both unordered

Yule condensed data of Sir Francis Galton (as reported by Karl Pearson 1900) on the eye-color of fathers and sons to the form given in exhibit 10. The logarithmic analysis given in panel C of that table shows -- exhibit 11 does so pictorially -- that the four cells along the diagonal where father and son have the same eye color contain two or three times as many cases as the stripe adjustments account for (log 2 = 0.30, log 3 = 0.48). Clearly this analysis has been effective in doing what it was asked to do. We are free, however, to ask whether what it was asked to do was the best choice.

The logarithmic analysis treats both factors -- father's eye color and son's eye color in exactly the same way -- it is truly symmetrical. For exhibits 1 to 7, which interrelate hair color with eye color in a single person, there seems to be no natural objection to such symmetry. In the present example this is not the case -- there are arguments against symmetry.

exhibit 10 of chapter 30 -- eye inheritance

Frequencies of different combinations of eye color
in father and son analyzed as two responses
1 = Blue; 2 = Blue green or Gray;
3 = Dark Gray or Hazel; 4 = Brown

A) COUNTS

Son's eye color	Father's eye color 1	2	3	4
1	194	70	41	30
2	83	124	41	36
3	25	34	55	23
4	56	36	43	109

B) LOGS of COUNTS -- diagonal entries underscored

1	2.29	1.85	1.61	1.48
2	1.92	2.09	1.61	1.56
3	1.40	1.53	1.74	1.36
4	1.75	1.56	1.63	2.04

C) ELEMENTARY ANALYSIS -- of logs, by midmean differences, outside bold

					eff	fit	half	(*)
1	.45	.04	-.07	-.11	.04	1.74	.89	7.7
2	.04	.24	-.11	-.07	.08	1.78	.93	8.5
3	-.17	-.01	.33	+.04	-.23	1.47	.62	4.2
4	.00	-.16	.04	.54	-.05	1.65	.80	6.3
eff	.10	.07	-.06	-.15	1.70	0.00	-.85	
fit	1.80	1.77	1.64	1.55	0.00	-1.70	.85	
half	.95	.92	.79	.70	-.85	.85	0.00	
(*)	8.9	8.3	6.1	5.0				

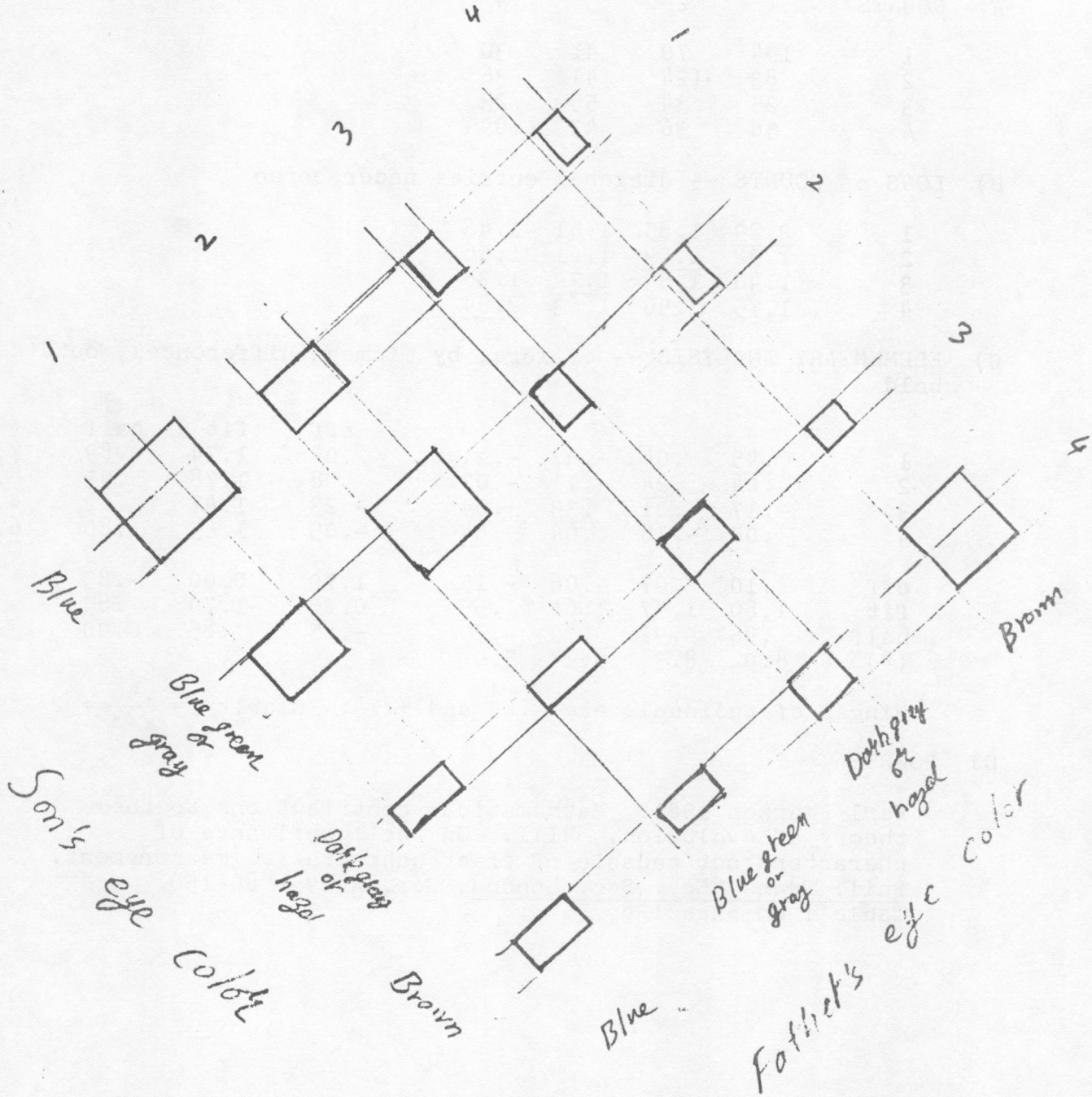
Hinges of residuals are -.09 and +.14. Clarity = $\frac{.56}{.23}$ = 2

D) SOURCE

Karl Pearson 1901? Mathematical contributions to the theory of evolution. VIII. On the inheritance of characters not capable of exact quantitative measurement. Phil. Trans. Roy. Soc. London, Ser. A 195: 80-150. Table I on page 138.

exhibit 11 of chapter 30 -- eye inheritance

Fitted (light rectangles) and observed (heavy rectangles) counts of father's and son's eye color (area scale, see exhibit 10 for fitting)



One strong argument can begin by pointing out that you cannot have sons without fathers, and go on to argue that we ought therefore to look at four distributions of son's eye color, one for each parent's eye color. If we accept this argument, the natural presentation involves a reference fit:

- obtained by combining (multiplying) the overall distribution of son's eye colors -- probably from panel C of exhibit 10 -- with the observed number of fathers of a given eye color.

Exhibit 12 shows the computations; exhibit 13 pictures the results. Again we see that agreement between father and son is more common than these reference values would suggest.

In this particular example, the argument that we should consider the data as four distributions -- one for each father's eye color -- was strong but not absolutely clinching. There is a counterargument, focused upon the fact that the sons are the fathers of the next generation, and arguing that a man cannot enter this table as a father unless he has a son.

No such argument can be applied to the data in exhibit 14, which reports blood group distributions for six tribes of Arizona indians. There is little sense in picking out an Rh pattern and asking how it is distributed over tribes.

exhibit 12 of chapter 30 -- eye inheritance

Calculations of suspensions for rootobars expressing the counts of panel A of exhibit 6 as unordered responses for an unordered characteristic.

A) BASE DISTRIBUTION

Son's eyes	Row effects (from panel C of exhibit 6)		
	as logs	as ratios	as fractions
1	.04	1.10	29.0%
2	.08	1.20	31.8%
3	-.23	.59	15.6%
4	-.05	.89	23.6%

B) FITTED COUNTS -- fractions from A applied to sums

	Father's eyes			
(Sum)	1 (358)	2 (264)	3 (180)	4 (198)
1	103.8	76.6	52.3	57.4
2	113.9	84.0	57.2	63.0
3	55.9	41.2	28.1	30.9
4	84.6	62.3	42.5	46.8

Example: 103.8 = 29.0% of 358, 113.9 = 31.8% of 358

C) FITTED DOUBLE ROOTS -- $\sqrt{4(\text{fit}) + 1}$

1	20.4	17.5	14.5	15.2
2	21.4	18.4	15.2	15.9
3	15.0	12.9	10.6	10.2
4	15.4	16.0	13.1	13.7

Example: $20.4 = \sqrt{4(103.8) + 1}$

(Cont'd)

essence of the matter. If fluctuations and irregularities had been so large -- as they often are -- that we would be lucky to have modest confidence in pointing out some cells as "high" and others as "low", a square root comparison -- after a logarithmic analysis -- may do better in getting to those few aspects of the matter that might then be reachable.

This contrast between the usefulnesses of square-root and logarithmic methods, when both are applicable, applies generally, although we did not emphasize this enough in chapter 24. Roots help us in dealing with irregularities and fluctuations. Logs help us once the data is massive enough to be decisive. (As a variety of long-tailed examples show, logs sometimes help us in dealing with greater irregularities and fluctuations.)

* logs are not all alike *

We need to be clear about how the use of logs for the Badeners differs from the use of logs for Neiswander's grub counts. With grubs we went to logs -- more specifically to log(count + free start) with a free start probably between 0.25 and 1 -- because this gave the simplest description of the data. In logs, row and column effects accounted for nearly everything that seemed to be going on with the grub counts. We were free to choose our form of expression -- we wanted to choose it to make description simple.

In dealing with the Badeners, a man who wants to use an elementary analysis feels forced to go to logs, whether or not the results are simple. Our attitude toward the world -- as summed up in our definition of unaffectedness -- demands that an analysis of paired responses into a SUM of terms should be an analysis of logs. If some other form of expression were simpler, we would not use it.

Consider the dreamed-up data of exhibit 9. If 1234 and ABC referred to circumstances, we would feel free to use any form of expression that would help. The analysis in terms of raw counts -- see panel B -- cannot be improved on. Under these circumstances we would seize it.

If 1234 and ABC referred to responses -- as in dealing with the Badeners -- we would insist on analyzing log counts. Reading panel D, we would have to say that row and column fits do not tell the whole story -- each response fails to be unaffected by the other. This is clearly the case. Compare columns 1 and 4:

- in column 1 there are 400/1800 = 22.2% A's.
- in column 4 there are 100/900 = 11.1% A's.

Clearly the ABC response is not unaffected by selection on the 1234 response.

If unaffectedness is our standard of reference, an elementary analysis of log counts is very much more to

exhibit 9 of chapter 30 -- casual example

A set of counted data -- and two analyses

A) The COUNTS -- obviously invented for the analysis

	1	2	3	4
A	400	300	200	100
B	600	500	400	300
C	800	700	600	500

B) ELEMENTARY ANALYSIS of RAW COUNT

	1	2	3	4	eff	fit
A	0	0	0	0	-200	250
B	0	0	0	0	0	450
C	0	0	0	0	200	650
eff	150	50	-5	-50	450	000
fit	600	500	400	300	000	-450

C) LOGS of COUNTS

	1	2	3	4
A	2.60	2.48	2.30	2.00
B	2.78	2.70	2.60	2.48
C	2.90	2.85	2.78	2.60

D) ELEMENTARY ANALYSIS of C

	1	2	3	4	eff	fit
A	.08	.04	-.02	-.12	-.12	2.47
B	-.03	-.03	-.01	.07	.07	2.66
C	-.05	-.02	.03	.05	.05	2.64
eff	.17	.09	-.03	-.23	2.59	0.00
fit	2.76	2.68	2.56	2.36	0.00	-2.59

the point than are elementary analysis of say, raw counts. This is generally true. In this particular example, if simple analysis is our aim, the opposite applies. There can be very real differences.

Not all elementary analyses of log counts are alike.

30E. Circumstance and response, both unordered

Yule condensed data of Sir Francis Galton (as reported by Karl Pearson 1900) on the eye-color of fathers and sons to the form given in exhibit 10. The logarithmic analysis given in panel C of that table shows -- exhibit 11 does so pictorially -- that the four cells along the diagonal where father and son have the same eye color contain two or three times as many cases as the stripe adjustments account for (log 2 = 0.30, log 3 = 0.48). Clearly this analysis has been effective in doing what it was asked to do. We are free, however, to ask whether what it was asked to do was the best choice.

The logarithmic analysis treats both factors -- father's eye color and son's eye color in exactly the same way -- it is truly symmetrical. For exhibits 1 to 7, which interrelate hair color with eye color in a single person, there seems to be no natural objection to such symmetry. In the present example this is not the case -- there are arguments against symmetry.

exhibit 10 of chapter 30 -- eye inheritance

Frequencies of different combinations of eye color
in father and son analyzed as two responses
1 = Blue; 2 = Blue green or Gray;
3 = Dark Gray or Hazel; 4 = Brown

A) COUNTS

Son's eye color	Father's eye color 1	2	3	4
1	194	70	41	30
2	83	124	41	36
3	25	34	55	23
4	56	36	43	109

B) LOGS of COUNTS -- diagonal entries underscored

1	2.29	1.85	1.61	1.48
2	1.92	2.09	1.61	1.56
3	1.40	1.53	1.74	1.36
4	1.75	1.56	1.63	2.04

C) ELEMENTARY ANALYSIS -- of logs, by midmean differences, outside bold

					eff	fit	half	(*)
1	.45	.04	-.07	-.11	.04	1.74	.89	7.7
2	.04	.24	-.11	-.07	.08	1.78	.93	8.5
3	-.17	-.01	.33	+.04	-.23	1.47	.62	4.2
4	.00	-.16	.04	.54	-.05	1.65	.80	6.3
eff	.10	.07	-.06	-.15	1.70	0.00	-.85	
fit	1.80	1.77	1.64	1.55	0.00	-1.70	.85	
half	.95	.92	.79	.70	-.85	.85	0.00	
(*)	8.9	8.3	6.1	5.0				

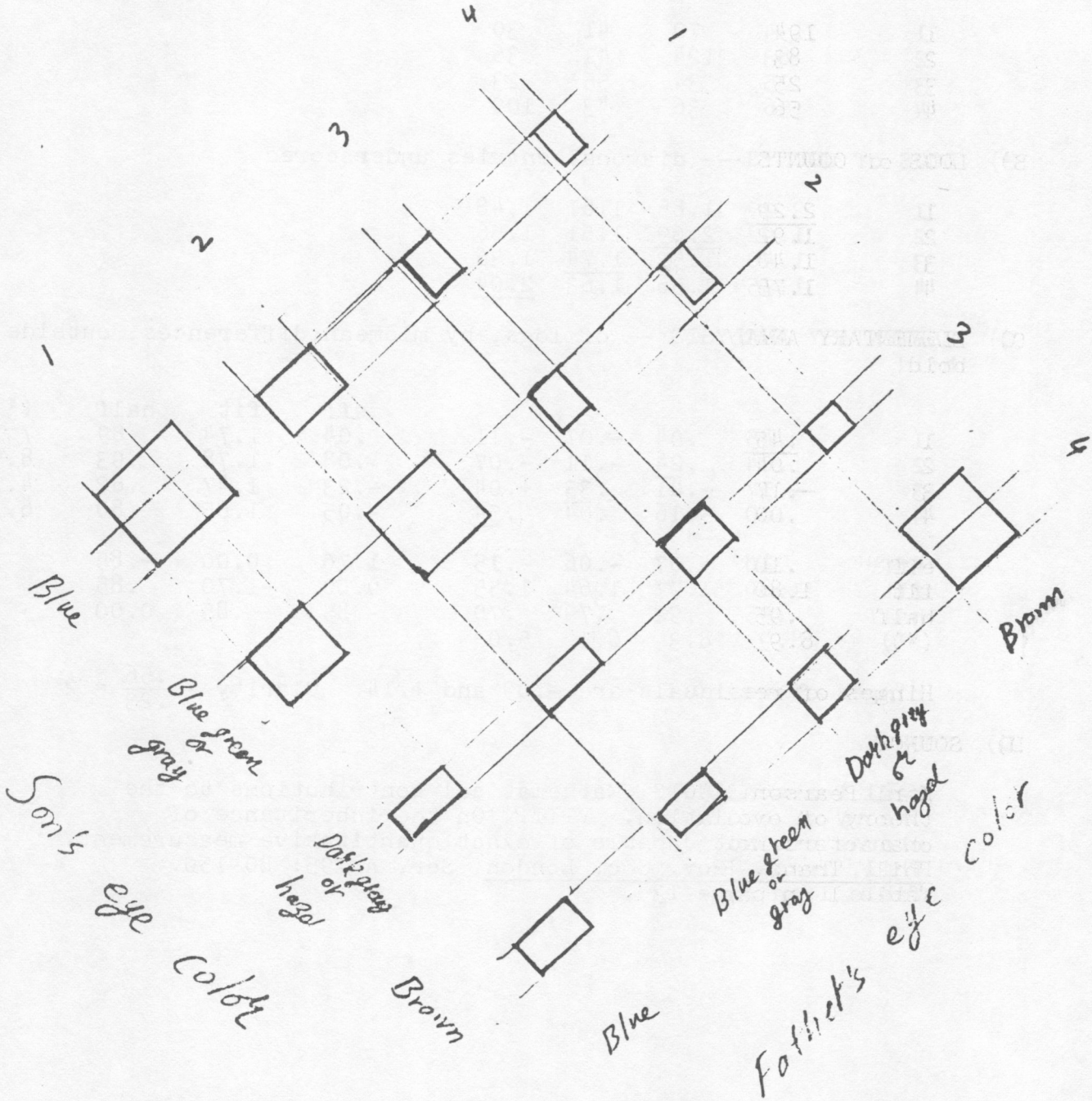
Hinges of residuals are -.09 and +.14. Clarity = $\frac{.56}{.23}$ = 2

D) SOURCE

Karl Pearson 1901? Mathematical contributions to the theory of evolution. VIII. On the inheritance of characters not capable of exact quantitative measurement. Phil. Trans. Roy. Soc. London, Ser. A 195: 80-150. Table I on page 138.

exhibit 11 of chapter 30 -- eye inheritance

Fitted (light rectangles) and observed (heavy rectangles) counts of father's and son's eye color (area scale, see exhibit 10 for fitting)



One strong argument can begin by pointing out that you cannot have sons without fathers, and go on to argue that we ought therefore to look at four distributions of son's eye color, one for each parent's eye color. If we accept this argument, the natural presentation involves a reference fit:

> -- obtained by combining (multiplying) the overall distribution of son's eye colors -- probably from panel C of exhibit 10 -- with the observed number of fathers of a given eye color.

Exhibit 12 shows the computations; exhibit 13 pictures the results. Again we see that agreement between father and son is more common than these reference values would suggest.

In this particular example, the argument that we should consider the data as four distributions -- one for each father's eye color -- was strong but not absolutely clinching. There is a counterargument, focused upon the fact that the sons are the fathers of the next generation, and arguing that a man cannot enter this table as a father unless he has a son.

No such argument can be applied to the data in exhibit 14, which reports blood group distributions for six tribes of Arizona indians. There is little sense in picking out an Rh pattern and asking how it is distributed over tribes.

exhibit 12 of chapter 30 -- eye inheritance

Calculations of suspensions for rootobars expressing the counts of panel A of exhibit 6 as unordered responses for an unordered characteristic.

A) BASE DISTRIBUTION

Son's eyes	Row effects (from panel C of exhibit 6)		
	as logs	as ratios	as fractions
1	.04	1.10	29.0%
2	.08	1.20	31.8%
3	-.23	.59	15.6%
4	-.05	.89	23.6%

B) FITTED COUNTS -- fractions from A applied to sums

	Father's eyes			
	1	2	3	4
(Sum)	(358)	(264)	(180)	(198)
1	103.8	76.6	52.3	57.4
2	113.9	84.0	57.2	63.0
3	55.9	41.2	28.1	30.9
4	84.6	62.3	42.5	46.8

Example: 103.8 = 29.0% of 358, 113.9 = 31.8% of 358

C) FITTED DOUBLE ROOTS -- $\sqrt{4(\text{fit}) + 1}$

1	20.4	17.5	14.5	15.2
2	21.4	18.4	15.2	15.9
3	15.0	12.9	10.6	10.2
4	15.4	16.0	13.1	13.7

Example: $20.4 = \sqrt{4(103.8) + 1}$

(Cont'd)

exhibit 12 (cont'd.)

D) OBSERVED DOUBLE ROOTS -- $\sqrt{4(\text{obs}) + 2}$

1	27.9	16.8	12.9	11.0
2	18.3	22.4	12.9	12.1
3	10.1	11.7	14.9	9.7
4	15.1	12.1	13.2	21.0

Example: $27.9 = \sqrt{4(194) + 2}$

E) RESIDUALS -- panel D minus panel C

1	7.5	-0.7	-1.6	-4.2
2	-3.1	4.0	-2.3	-3.6
3	-4.9	-1.2	4.3	-0.5
4	-0.3	-3.9	0.1	7.3

Example: 7.5 = 27.9 - 20.4

Hinges of residuals are -3.3h and +2.1

exhibit 13 of chapter 30 -- eye inheritance

Display of the fit of exhibit 12, based on applying
an overall distribution of son's eye color
to the actual number of fathers
(see exhibits 10 to 12)

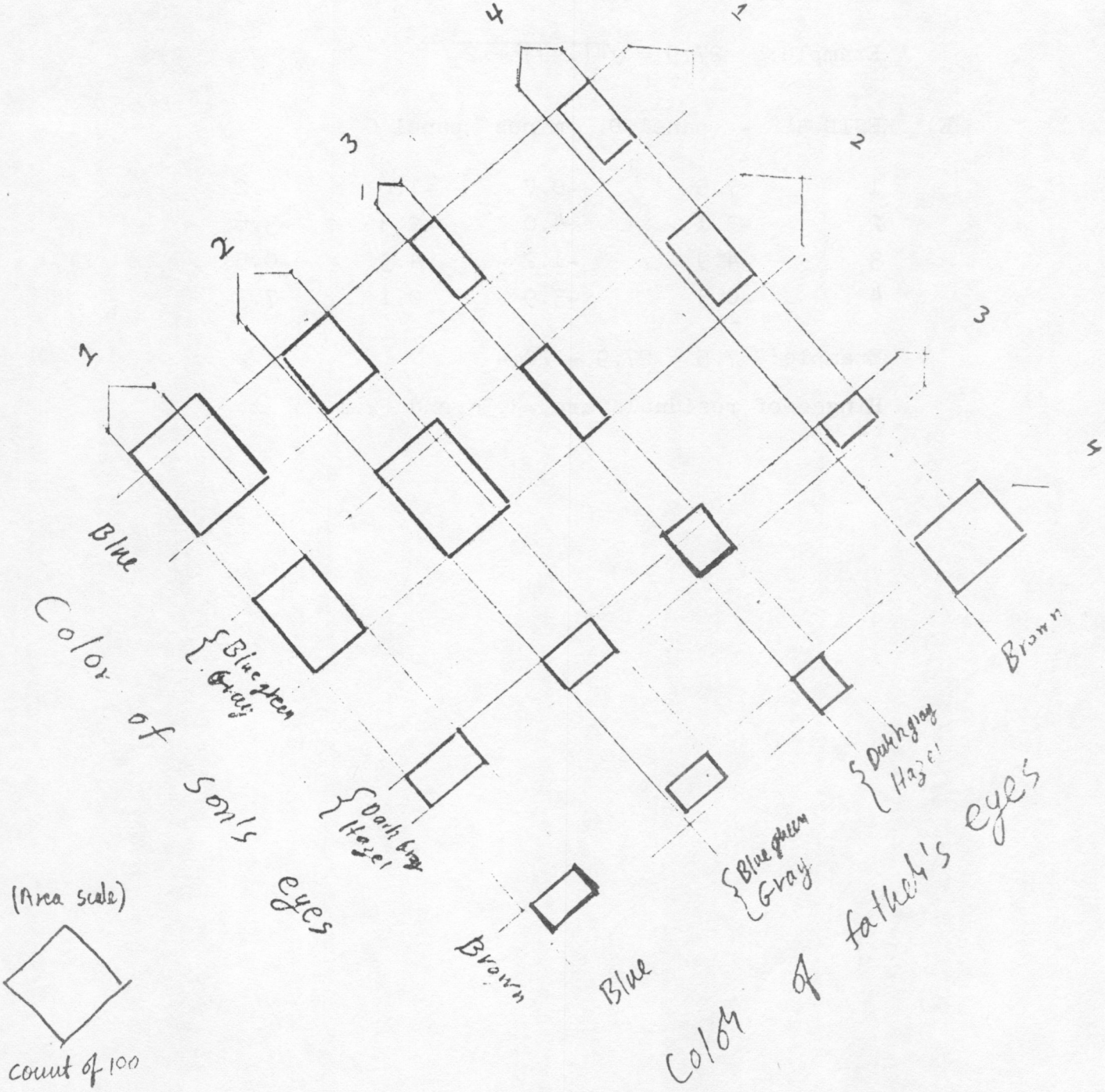


exhibit 14 of chapter 30 -- tribal Rh

Rh patterns for six Arizona tribes

A) PATTERNS -- reaction = +, nonreaction = -, not tested = blank

Pattern number	Rh-antiserum				
	rh′	hr′	Rh′	rh″	(combines)
1	+	-	+	-	
2	+	+	+	-	
3	-	+	+	+	
4	+	-	+	+	
5	+	+	+	+	
6	-	+	+	-	
7	+		+	+	(4 and 5)
8	+		+	-	(1 and 2)

B) COUNTS

Tribe	Pattern								(sum)
	1	2	3	4	5	6	7	8	
West Navaho	24	14	29	1	15	4	15	2	(104)
Apache*	47	22	64	2	40	3	0	0	(178)
Pima	136	59	104	20	139	6	22	3	(489)
Yuma	63	32	34	4	39	7	0	0	(179)
Hopi & Tewa	20	23	35	0	7	21	6	5	(117)
Maricopa	12	3	10	0	8	0	5	0	(38)

*Apache San Carlos

C) SOURCE K. S. Brown, B. L. Hanna, A. A, Dahlberg, H. H. Strandskov 1958. The distribution of blood group alleles among indians of southwest North America. Amer. J. Human Genetics 10: 175-195. Table 1 on page 177 .

D) PROBLEM Carry through the analogs of exhibits 10 and 12 for this data.

(How large a sample do you want to take from each tribe?) There is much sense in picking a tribe and asking how its members are distributed over Rh patterns.

So long as there are neither too many rows nor too many columns, we can try the approach we have used in this section whenever it seems appropriate. If there are too many versions of either the circumstances or the response for our arithmetic, however, we face a difficult situation.

If the opposite happens and some of the response versions are very popular, we face a puzzling question. In dealing with counts for a response, we are really dealing with counted fractions. Yet we have treated our numbers as if they were pure counts, not numerators of counted fractions. So long as all our cells are small -- so long as their counts are only small fractions of the total count -- this is obviously sound. For if we were to treat them as counted fractions, we would find the f-roots of the counts almost exactly a straight-line function of the roots of the counts -- or the f-logs almost exactly proportional to the logs of the counts.

But what if there are some large cells -- some cells so large as to make working with counted fractions different in result from working with counts? Which should we use? Square roots or f-roots? Lots or f-logs?

And the echo comes back "which?". The best advice so far available is very general:

- when in doubt, try both -- and use the one that gives the cleanest analysis.

*ordering can sometimes be extracted *

There are situations where a useful ordering of responses can be found, even though it was not given in advance. Sometimes we can find such orderings by merely looking hard at the data.

Examples where an order can be extracted are probably more common than they appear to be. Data of this sort tends, we may suspect, to be reported less often than data which can be more directly analyzed.

In many cases, we can look at the counts and see what to do. Exhibit 15 gives the counts for an example given by Zeisel concerning women's favorite remedies for various ailments. Rearrangement into the orderings of panel C is a matter of:

- neglecting "others."
- looking at the blank cells.
- trying to fill in between.

The result is fairly clear. (We will se a more systematic treatment when we come to chapter 30x.)

exhibit 15 of chapter 30 -- favorite remedies

Women's favorite remedies for various ailments

A) DATA

Ailment	Favorite Remedy				
	AA	BB	CC	other	(sum)
Neuralgia	104	47	198	24	(376)
Cold	98	41	401	30	(570)
Sour Stomach	302	60	-	23	(385)
Headache	-	242	210	26	(478)
(sum)	(507)	(390)	(809)	(103)	(1809)

B) SOURCE

H. Zeisel 1947. Say it with figures

Table V-7 on page 93.

C) DATA REARRANGED NEATLY

	Favorite Remedy		
	AA	BB	CC
Sour Stomach	302	60	-
Neuralgia	104	47	198
Cold	98	41	401
Headache	-	242	210

30F. Ordered circumstance and unordered response

Unordered responses can occur with any kind of second factor. One case we have not yet looked at is an ordered circumstance. For this, we may turn to counts given by Thieme (1952) for the distribution of OAB blood groups in Puerto Rico. Thieme classified all this subjects (by combining "skin color," "hair form," and "lip thickness and nose shape") on a scale from 0 (most typically white) to 9 (most typically negro) and gave blood-group distributions for each class. Only six individuals were "0" and only one individual was "9", so these two classes have been omitted from exhibit 16.

In looking into this data we are concerned with how the distribution of blood groups shifts as we move along the ethnic scale. Since the fraction in each of the first four blood groups is relatively large, we can approach the analysis of this data by examining how each of these four fractions vary as we move along that scale. It again seems natural and reasonably to express these fractions as f-roots. Having done this, we can expose the results to view as in exhibit 17. We see here that:

- the frequency of Group O appears to trend slowly upward.

- that of Group A_1, after an uncertain start, trends firmly downward.

exhibit 16 of chapter 30 -- blood groups

Thieme's blood group distributions for eight levels of ethnic appearance in Puerto Rico

A) COUNTS

Ethnic Appear.	Blood group O	A_1	A_2	B	A_1B	A_2B	(sum)
1	50	27	11	10	1	2	(101)
2	450	214	88	63	9	4	(828)
3	787	382	136	112	8	16	(1441)
4	224	109	40	50	6	4	(433)
5	70	26	9	22	5	1	(133)
6	89	28	14	24	1	3	(159)
7	54	15	10	18	1	3	(101)
8	23	4	3	10	1	1	(42)
							((3238))

B) PERCENTS -- among those of given ethnic appearance

1	49.5	26.7	10.9	9.9	-	-
2	54.3	25.8	10.6	7.6	-	-
3	54.6	26.5	9.4	7.8	-	-
4	51.7	25.2	9.2	11.6	-	-
5	52.6	19.6	6.8	16.5	-	-
6	56.0	17.6	8.8	15.1	-	-
7	53.5	14.8	9.9	17.8	-	-
8	54.8	9.5	7.1	23.8	-	-

C) F-ROOTS -- corresponding to percents

1	-.01	-.48	-.87	-.89	-	-
2	.09	-.50	-.88	-.97	-	-
3	.09	-.48	-.91	-.96	-	-
4	.03	-.52	-.92	-.85	-	-
5	.05	-.63	-1.00	-.72	-	-
6	.12	-.69	-.93	-.77	-	-
7	.07	-.77	-.89	-.68	-	-
8	.10	-.91	-.99	-.54	-	-

D) MIDMEANS END THIRDS -- raw thirds.

1,2,3	.07	-.49	-.89	-.94	-	-
6,7,8	.10	-.78	-.93	-.67	-	-

E) COUNTS -- for approximately one third of cases.

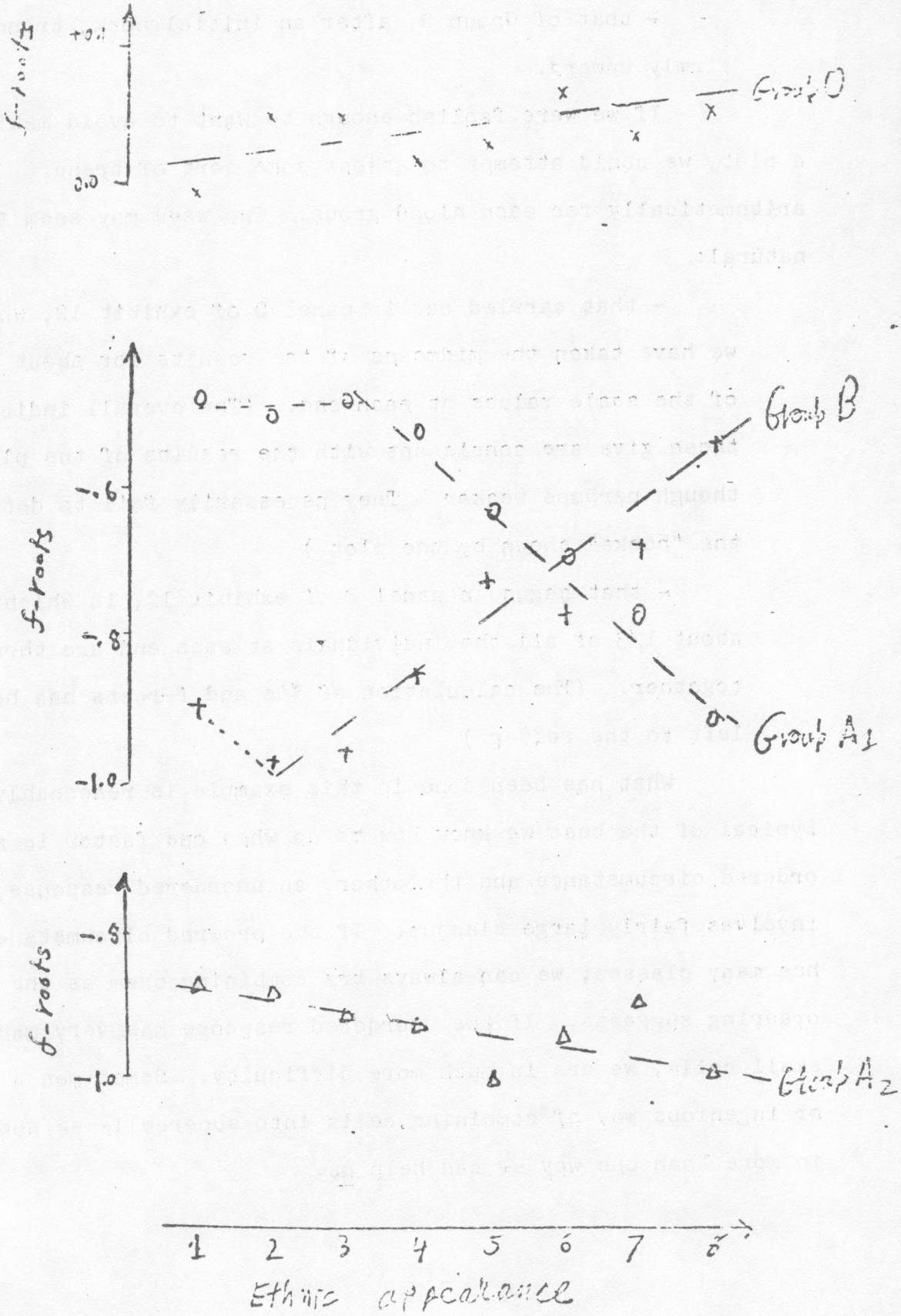
1,2	500	241	99	73	(10)	(6)	(929)
4,5,6,7,8	460	182	76	124	(14)	(12)	(868)

(further steps left to reader)

SOURCE: F. P. Thieme 1952. The geographic and racial distribution of ABO and Rh blood types and tasters of PTC in Puerto Rico. Amer. J. Juman Genetics 4: 94-112. Table 3 on page 101.

exhibit 17 of chapter 30 -- blood groups

Dependence of blood group fractions on ethnic appearance



- that of Group A_2, trends slowly downward.

- that of Group B, after an initial hook, trends firmly upward.

If we were foolish enough to want to avoid making a plot, we could attempt to assess some sort of trend arithmetically for each blood group. Two ways may seem fairly natural:

- that carried out in panel D of exhibit 12, where we have taken the midmeans of the results for about 1/3 of the scale values at each end. (The overall indications these give are consistent with the results of the plot, though perhaps weaker. They necessarily fail to detect the "hooks" shown by the plot.)

- that begun in panel E of exhibit 12, in which about 1/3 of all the individuals at each end are thrown together. (The calculation of %'s and f-roots has been left to the reader.)

What has been done in this example is reasonably typical of the best we know how to do when one factor is an ordered circumstance and the other, an unordered response, involves fairly large classes. If the ordered circumstance has many classes, we can always try combining them as the ordering suggests. If the unordered response has very many small cells, we are in much more difficulty. Sometimes a wise or ingenious way of combining cells into supercells -- possible in more than one way -- can help us.

* a second ordering? *

It is often rather easy to order the response, given an order for the circumstances. Exhibit 18 shows the % of stems made up of trees of certain species at 13 composite stations, all at altitudes between 2500 and 3500 feet in the Great Smoky Mountains. The stations are numbered in order of decreasing humidity from wet to dry. Clearly the order of species is useful, though one might argue for a modification if this were all the data. (The order adopted by Whittaker was based on analyses for three elevation bands, not just this one.)

exhibit 18 of chapter 30 -- tree prevalence

Prevalences of different species at composite stations in the Great Smoky Mountains.

(Altitude 2500 to 3500 feet, %'s are for the 1-inch size class and larger.)

A) DATA

Species	% at composite station						
	1	3	5	7	9	11	13
Tilia heterophylla	17	14	1				
Acer saccharum	6	14	1				
Tsuga canadensis	25	6	9				
Betula allegheniensis	26	13	5				
Liriodendron tulipifem	2	1	10	1			
Halesia monticola	5	28	3	1	1		
Acer rubrum	P	2	27	35	16	7	3
Hamamelis virginiana		1	6	9			
Castanea dentata (dead)		2	1	10	15	5	
Oxydendrum arboreum			5	13	23	9	2
Quercus coccinea				1	1	15	10
Pinus rigida					2	20	40
Pinus pungens					5	38	41

B) SOURCE

R. H. Whittaker 1956. Vegetation of the Great Smoky Mountains. Ecological Monographs 26:1-80. Table 2 on page 7, omitting even-numbered stations and all species for which no prevalence reaches 10%. (Composite stations grouped by weighted averages of stems by class.)

****

30H. <u>How far have we come</u>?

If a two-way table conceals two responses, the natural basis of comparison is one of "unaffectedness". If knowledge of one response tells us precisely nothing new about the other, we have a relation

cell count = (row number)·(column number).

This becomes an additive relation -- easy to handle -- if we take logs of all counts. Elementary analysis by midmean differences gives us -- after calculation -- a fitted count for each cell. Comparison of observed and fitted counts by differences of double roots then gives us an idea of what seems to be going on.

When one factor is an unordered circumstance and the other an unordered response, we tend to compare each cell count with a fitted value of the form

(total count for the circumstance)

times

(pooled fraction for the response).

If we have an ordered circumstance, and an unordered response in not too many versions, we find analysis considerably simpler, since we now have chance to see trends and the like.

Finally, there are a variety of situations where we can extract orderings from the data. Two simple examples have been given, others of greater interest will appear in chapter 30x.

*****

30K. A further example

In 1906, Karl Pearson published the table with which exhibit 19 opens. Both girl's intelligence and character of hair are clearly to be regarded as responses. One is ordered, the other, at least on the face of matters, is not. The natural first step is to convert the counts along the ordered concealed response into f-logs. Once this is done, the proper reordering of character of hair is clear. The reader is encouraged to consider why wavy hair seems to go with higher intelligence and curly hair with lower.

exhibit 19 of chapter 30 -- hair and intelligence

Character of girl's hair and intelligence

A) TAGS

QI = Quick Intelligent
I = Intelligent
SI = Slow Intelligent
S = Slow
SD = Slow Dull
VS = Very Dull

B) COUNTS

Girl's hair	Girl's intelligence QI	I	SI	S	SD	VD	(sum)
Smooth	198h	460	415	174h	72	25	(1345)
Wavy	90	80	144h	52h	20	6	(493)
Curly	30	76	62	37	19	7	(231)
(sum)	(318h)	(716)	(621h)	(264)	(111)	(38)	2069

C) f-LOGS -- unit = 0.01

	Cut in intelligence (Between tags indicated) QI/I	I/SI	SI/S	S/SD	SD/VD
Smooth	-88	-2	68	129	200
Wavy	-76	10	84	147	221
Curly	-98	-12	41	105	176

D) Same REARRANGED

						mean
Curly	-98	-12	41	105	176	
Smooth	-88	-2	68	129	200	
Wavy	-76	10	84	147	221	
S - C	10	10	27	24	24	19
W - S	12	12	16	18	21	16

(Cont'd)

exhibit 19 (cont'd.)

E) The corresponding ELEMENTARY ANALYSIS -- unit = 0.01

	cut in intelligence						
Hair	QI/I	I/SI	SI/S	S/SD	SD/VD	eff	fit
Curly	7	7	-5	-4	-5	-19	42
Smooth	-2	-2	3	1	0	0	61
Wavy	-6	-6	3	3	5	16	71
eff	-147	-61	4	67	139	61	00
fit	-86	0	65	128	200	00	-61

F) PROBLEM

Make an extended elementary analysis.

G) SOURCE

K. Pearson 1906. On the relationship of intelligence to size and shape of head, and to other physical and neutral characters. Biometrika 5: 105-146. Table XXX on page 143.

****

30P. Problems for chapter 30

1. Do the problem of exhibit 14.

2. Do the problem of exhibit 19.

3. Omitting patterns 7 and 8, which come from incomplete testing, and Maricopas, who are few in number, the counts of exhibit 9 produce the % of exhibit 20. Complete the table. Use common sense to rearrange both the patterns and the tribes. Give your result. Justify in some detail the order of the Pima and Yuma in comparison with the others.

4. Another example cited by Zeisel involves favorite radio programs at various economic levels. Exhibit 21 has the counts. Convert to % within income class. Rearrange types of programs in relation to what you then see. Comment.

exhibit 20 of chapter 30

Rh patterns ready for ordination

(see exhibit 9)

A) PERCENT OF PATTERNS IN TRIBES

Tribe	Pattern 1	2	3	4	5	6	(total count
West Navaho	28	10	33	1	17	6	(87)
Apache San Carlos	26	23	36	1	22	2	(178)
Pima	29	13	22	4	30	1	(464)
Yuma	35	18		2		4	(179)
Hopi & Tewa	19		33	0	7		(106)

exhibit 21 of chapter 30

Favorite radio programs at various economic levels

A) COUNTS

Type of program	Income class A	B	C	D	(Sum)
Dance & others	107	331	443	201	(1082)
Variety	85	288	432	202	(1007)
Serials	33	108	175	75	(391)
News	44	90	87	32	(251)
Drama	15	66	79	30	(190)
Classical Music	25	65	66	6	(162)
Religious	4	15	23	11	(53)
(Sum)	(313)	(963)	(1305)	(557)	3138

B) SOURCE

H. Zeisel 1947. Say it with Figures. Table V-4 on page 91.